Network Working Group
Request for Comments: 1154

D. Robinson
R. Ullmann
Prime Computer, Inc.
April 1990

Encoding Header Field for Internet Messages

1. Status of the Memo

This RFC proposes an elective experimental Encoding header field to permit the mailing of multi-part, multi-structured messages.

The use of Encoding updates RFC 1049 (Content-Type), and is a suggested update to RFCs 1113, 1114, and 1115 (Privacy Enhancement) [4,7,8].

Distribution of this memo is unlimited.

2. Introduction

RFC 822 [2] defines an electronic mail message to consist of two parts, the message header and the message body, separated by an apparently blank line.

The Encoding header field permits the message body itself to be further broken up into parts, each part also separated from the next by an apparently blank line.

Thus, conceptually, a message has a header part, followed by one or more body parts, all separated by blank lines.

Each body part has an encoding type. The default (no Encoding field in the header) is a message body of one part of type "text".

3. The Encoding Field

The Encoding field consists of one or more subfields, separated by commas. Each subfield corresponds to a part of the message, in the order of that part's appearance. A subfield consists of a line count, a keyword defining the encoding, and optional information relevant only to the specific encoding. The line count is optional in the last subfield.

3.1. Format of the Encoding Field

The format of the Encoding field is:

[<count> <keyword> [<options>],]* [<count>] <keyword> [<options>]

where:

<count> := a decimal integer
<keyword> := a single alphanumeric token starting with an alpha
<options> := keyword-dependent options

3.2. <count>

The line count is a decimal number specifying the number of text lines in the part. Parts are separated by a blank line, which is not included in the count of either the proceeding or following part. Because a count always begins with a digit and a keywords always begins with an letter, it is always possible to determine if the count is present. (The count is first because it is the only information of interest when skipping over the part.)

The count is not required on the last or only part.

3.3. <keyword>

The keyword defines the encoding type. The keyword is a common single word name for the encoding type. The keywords are not case-sensitive.

The list of standard keywords is intended to be the same as the list used for the Content-Type: header described in [6]. This RFC proposes additions to the list. Implementations can then treat "Content-Type" as an alias of "Encoding", which will always have only one body part.

3.4. <options>

The optional information is used to specify additional keyword-specific information needed for interpreting the contents of the encoded part. It is any sequence of tokens not containing a comma.

3.5. Encoding Version Numbers

In general, version numbers for encodings, when not actually available within the contents of the encoded information, will be handled as options.

3.6. Comments

Comments enclosed in parentheses may, of course, be inserted anywhere in the Encoding field. Mail reading systems may pass the comments to

their clients. Comments must not be used by mail reading systems for content interpretation; that is the function of options.

4. Encodings

This section describes some of the defined encodings used.

As with the other keyword-defined parts of the header format standard, extensions in the form of new keywords are expected and welcomed. Several basic principles should be followed in adding encodings:

- The keyword should be the most common single word name for the encoding, including acronyms if appropriate. The intent is that different implementors will be likely to choose the same name for the same encoding.

- Keywords not be too general: "binary" would have been a bad choice for the "hex" encoding.

- The encoding should be as free from unnecessary idiosyncracies as possible, except when conforming to an existing standard, in which case there is nothing that can be done.

- The encoding should, if possible, use only the 7 bit ASCII printing characters if it is a complete transformation of a source document (e.g., "hex" or "uuencode"). If it is essentially a text format, the full range may be used. If there is an external standard, the character set may already be defined.

Keywords beginning with "X-" are permanently reserved to implementation-specific use. No standard registered encoding keyword will ever begin with "X-".

4.1. Text

This indicates that the message is in no particular encoded format, but is to be presented to the user as is.

The full range of the ASCII character set is used. The message is expected to consist of lines of reasonable length (less than 1000 characters).

On some transport services, only the 7 bit subset of ASCII can be used. Where full 8 bit transparency is available, the text is assumed to be ISO 8859-1 [3] (ASCII-8).

4.2. Message

This encoding indicates that the body part is itself in the format of an Internet message, with its own header part and body part(s). A "message" body part's message header may be a full internet message header or it may consist only of an Encoding field.

Using the message encoding on returned mail makes it practical for a mail reading system to implement a reliable resending function, if the mailer generates it when returning contents. It is also useful in a "copy append" MUA operation.

Message encoding is also used when mapping to X.400 to handle recursively included X.400 P2 messages.

4.3. Hex

The encoding indicates that the body part contains binary data, encoded as 2 hexadecimal digits per byte, highest significant nibble first.

Lines consist of an even number of hexadecimal digits. Blank lines are not permitted. The decode process must accept lines with between 2 and 1000 characters, inclusive.

4.4. EVFU

EVFU (Electronic Vertical Format Unit) specifies that each line begins with a one-character "channel selector". The original purpose was to select a channel on a paper tape loop controlling the printer.

This encoding is sometimes called "FORTRAN" format. It is the default output format of FORTRAN programs on a number of computer systems.

The legal characters are '0' to '9', '+', '-', and space. These correspond to the 12 rows (and absence of a punch) on a printer control tape (used when the control unit was electromechanical).

The channels that have generally agreed definitions are:

1	advances to the first print line on the next page
0	skip a line, i.e., double-space
+	over-print the preceeding line
-	skip 2 lines, i.e., triple-space
(space)	print on the next line, single-space

4.5. EDI

The EDI (Electronic Document Interchange) keyword indicates that the message or part is a business document, formatted according to ANSI X12 or related standards.

The first word after the EDI keyword indicates the particular interchange standard.

A message containing a note and 2 X12 purchase orders might have an encoding of:

Encoding: 17 TEXT, 146 EDI X12, 69 EDI X12

4.6. X.400

The Encoding header field provides a mechanism for mapping multi-part messages between CCITT X.400 [1] and RFC 822.

The X.400 keyword specifies a section that is converted from an X.400 body part type not known to the gateway, or not corresponding to a useful internet encoding.

If the message transits another gate, or if the receiving user has the appropriate software, it can be decoded and used.

The X.400 keyword is followed by a second token indicating the method used. The simplest form is "X.400 HEX", with the complete X.409 encoding of the body part in hexadecimal. More compact is "X.400 3/4", using the 3-byte to 4-character encoding as specified in RFC 1113, section 4.3.2.4.

4.7. uuencode

The uuencode keyword specifies a section consisting of the output of the uuencode program supplied as part of uucp.

4.8. encrypted

The encrypted keyword indicates that the section is encrypted with the methods in RFC 1115 [8]. This replaces the possible use of RFC 934 [5] encapsulation.

References

[1] International Telegraph and Telephone Consultative Committee, "Data Communication Networks: Message Handling Systems", In CCITT Recommendations X.400 to X.430, VIIIth Plenary Assembly, Malaga-

Torremolinos, 1984, Fascicle VIII.7 ("Red Book").

[2] Crocker, D., "Standard for the Format of ARPA Internet Text Messages", RFC 822, University of Delaware, August 1982.

[3] International Organization for Standardization, "Information processing - 8-bit single-byte coded graphic character sets - Part 1: Latin alphabet No. 1", ISO 8859-1, ISO, 1987.

[4] Linn, J., "Privacy Enhancement for Internet Electronic Mail: Part I -- Message Encipherment and Authentication Procedures", RFC 1113, IAB Privacy Task Force, August 1989.

[5] Rose, M., and E. Stefferud, "Proposed Standard for Message Encapsulation", RFC 943, University of Delaware and NMA, January 1985.

[6] Sirbu, M., "Content-type Header Field for Internet Messages", RFC 1049, CMU, March 1988.

[7] Kent, S., and J. Linn, "Privacy Enhancement for Internet Electronic Mail: Part II -- Certificate-Based Key Management", RFC 1114, IAB Privacy Task Force, August 1989.

[8] Linn, J., "Privacy Enhancement for Internet Electronic Mail: Part III -- Algorithms, Modes, and Identifiers", RFC 1115, IAB Privacy Task Force, August 1989.

Security Considerations

Security issues are not addressed in this memo.

Authors' Addresses

David Robinson 10-30
Prime Computer, Inc.
500 Old Connecticut Path
Framingham, MA 01701

Phone: +1 508 879 2960 x1774

Email: DRB@Relay.Prime.COM

Robert Ullmann 10-30
Prime Computer, Inc.
500 Old Connecticut Path
Framingham, MA 01701

Phone: +1 508 879 2960 x1736

Email: Ariel@Relay.Prime.COM